Virus Research 211 (2016) 103–116

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Genetic structure of populations of sugarcane streak mosaic virus in China: Comparison with the populations in India

Zhen He[a,b,c], Ryosuke Yasaka[c,d], Wenfeng Li[e], Shifang Li[a,**], Kazusato Ohshima[c,d,*]

[a] State Key Laboratory of Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Yuanmingyuan West Road No. 2, Haidian District, Beijing 100193, PR China
[b] Department of Plant Pathology, School of Horticulture and Plant Protection, Yangzhou University, Wenhui East Road No. 48, Yangzhou, 225009 Jiangsu Province, PR China
[c] Laboratory of Plant Virology, Department of Applied Biological Sciences, Faculty of Agriculture, Saga University, 1-banchi, Honjo-machi, Saga 840-8502, Japan
[d] The United Graduate School of Agricultural Sciences, Kagoshima University, 1-21-24, Kagoshima 890-0065, Japan
[e] Yunnan Key Laboratory of Genetic Improvement of Sugarcane, Sugarcane Research Institute, Yunnan Academy of Agricultural Sciences, Lingquan East Road No. 363, Kaiyuan, 661600 Yunnan Province, PR China

ARTICLE INFO

Article history:
Received 10 August 2015
Received in revised form
24 September 2015
Accepted 25 September 2015
Available online 30 September 2015

Keywords:
Sugarcane streak mosaic virus
Poacevirus
Mixed infection
Population structure
Gene flow

ABSTRACT

Sugarcane streak mosaic virus (SCSMV) causes mosaic and streak symptoms on sugarcane and sorghum crops, and has a broad host range. SCSMV is a member of the genus *Poacevirus* in the family *Potyviridae*.Ten SCSMV isolates were collected from sugarcane plants showing mosaic and streaking in Southern China from 2009–2011. Sequence-based phylogenetic and population genetic analyses were conducted using four partial genomic sequences covering the full genomes. These analyses were used to estimate the sub-population differentiation and divergence within the Chinese virus population, and were compared with isolates from India. SCSMV-infected sugarcane plants in the field commonly harbor virus quasispecies (mutant cloud), and often have mixed infections with the same virus isolates. Inter- and intra-lineage recombination sites were identified in the protein 1, helper-component proteinase, coat protein and 3′ non-coding regions of the Chinese isolates. All the Chinese non-recombinant isolates fell into at least nine lineages, and many clustered with Indian isolates. However, estimates of genetic differentiation and gene flow indicated that the SCSMV populations in China and India are genetically independent. Our genetic study of a poacevirus population in South Asia regions indicates the importance of the evolutionary-based design to control viruses.

1. Introduction

Molecular evolution studies help us to understand the important features of RNA viruses, such as the molecular bases of virus geographical range, epidemiological routes, population structure, adaptation to new hosts, and the underlying evolutionary mechanisms. Therefore, these studies provide useful tools for designing better epidemic control strategies (Gibbs and Ohshima, 2010). Recently, several reports have been published on the genetic structure of viruses in the family *Potyviridae*; examples are potato virus Y (PVY) (Ogawa et al., 2008; Cuevas et al., 2012), soybean mosaic virus (SMV) (Seo et al., 2009), tobacco vein banding mosaic virus (TVBMV) (Zhang et al., 2011), zucchini yellow mosaic virus (ZYMV) (Lecoq et al., 2009), and turnip mosaic virus (TuMV) (Ohshima et al., 2002; Nguyen et al., 2013). These reports show that plant RNA virus populations have been shaped by selection, founder effects, and genetic recombination.

Sugarcane streak mosaic virus (SCSMV) is responsible for a mosaic disease that is widespread in sugarcane (*Saccharum officinarum*) and sorghum (*Sorghum* spp.). The disease was first identified by Hall et al. (1998) from quarantined sugarcane germplasm imported from Pakistan into the USA. SCSMV is a member of the new genus *Poacevirus* in the family *Potyviridae*, with triticum mosaic virus (TriMV) as its type species (Tatineni et al., 2009). SCSMV can be transmitted mechanically, although no obvious insect vectors have been identified to date (King

* Corresponding author at: Laboratory of Plant Virology, Department of Applied Biological Sciences, Faculty of Agriculture, Saga University, 1-banchi, Honjo-machi, Saga 840-8502, Japan. Fax: +81 952 28 8709.
** Corresponding author.
E-mail addresses: hezheng225@163.com, hezhen@yzu.edu.cn (Z. He), 14976004@edu.cc.saga-u.ac.jp (R. Yasaka), ynlwf@163.com (W. Li), sfli@ippcaas.cn (S. Li), ohshimak@cc.saga-u.ac.jp (K. Ohshima).

http://dx.doi.org/10.1016/j.virusres.2015.09.020
0168-1702/

et al., 2012). Poaceviruses consist of flexuous filamentous particles 680–750 nm long and 12–15 nm in diameter, and have a single-stranded, positive-sense RNA genome of about 10,000 nucleotides (nts) encoding a large polyprotein that is cleaved into ten mature proteins after translation by three virus-encoded proteinases (Fellers et al., 2009; Tatineni et al., 2009). Similar to the members of the genus *Potyvirus* (Chung et al., 2008), an additional protein (pretty interesting *Potyviridae* ORF, PIPO) was also predicted to be expressed as a fusion with the N-terminal region of the protein 3 (P3) in *Poacevirus* (Fellers et al., 2009). To date, SCSMV has a broad host range and has been found from sugarcane in India, Pakistan, USA, China, Thailand, Sri Lanka, Vietnam, and Indonesia (Hall et al., 1998; Chatenet et al., 2005; Damayanti and Putra, 2011; Li et al., 2011), from sorghum in India (Srinivas et al., 2010) and weeds in Indonesia (Prabowo et al., 2014). Seven full length genomic sequences of SCSMV isolates from Pakistan, Thailand, India, and China have been published (Xu et al., 2010; Li et al., 2011; Parameswari et al., 2013). The genomes of these isolates were 9724–9786 nt in length and had single open reading frames which encode predicted polyproteins of 3131–3133 amino acids.

Earlier studies have shown the evolutionary relationships and the molecular variability of SCSMV isolates from India (Viswanathan et al., 2008; Bagyalakshmi et al., 2012; Parameswari et al., 2013, 2014) and China (He et al., 2014) using only partial genomic sequences of one or two protein coding regions. In the present study, we extend our evolutionary investigation by using full genomic sequences to assess accurate phylogenetic population structures for SCSMV in South Asia. Our analyses, to our knowledge, provide the first depth definition of the phylogeographical structure of a poacevirus population.

2. Material and methods

2.1. Virus isolates

Sugarcane leaves showing mosaic symptom were collected in sugarcane growing areas of southern China (mostly from Yunnan province) during 2009 and 2011. Some isolates were imported from foreign countries around 30 years ago and maintained in the field at The National Nursery for Sugarcane Germplasm Resources, Yunnan, China. SCSMV infection of fresh leaf samples was assayed by reverse transcription and polymerase chain reaction (RT-PCR) using SCSMV specific primers in the coat protein (CP) coding region (data not shown), and the infected leaves were stored at −80 °C until use. Details of the Chinese SCSMV isolates, their place of origin, year of collection, original sugarcane variety, and symptoms are shown in Table 1, together with details of the isolates used in the analyses that were retrieved from International nucleotide sequence databases (Table S1).

2.2. Viral RNA and nucleotide sequencing

Viral RNA was extracted from SCSMV-infected sugarcane leaves using Isogen (Nippon Gene, Japan). We used a PrimeScript™ II High Fidelity One Step RT-PCR kit (Takara Bio, Japan) for cDNA synthesis and PCR. The kit contains PrimeSTAR® GXL DNA Polymerase. The RT-PCR products were separated by electrophoresis in agarose gels and purified using the QIA quick Gel Extraction Kit (Qiagen K.K., Japan). The resulting fragments were cloned into the pZErO-2 vector (Invitrogen, Japan). Full genomic sequences of each isolate were determined using four overlapping RT-PCR products; the 5′ non-coding region (NCR) excluding 27 nt of primer sequence used for RT-PCR to the middle of the protein 3 (P3) coding region (5NCRP3, 3226 nt, corresponding to nucleotides 28–3253 of SCSMV ID isolate genome), the N-terminal region of helper-component proteinase (HC-Pro) to the middle of the 6 kDa 2 (6K2) protein coding regions (HC6K2, 3156 nt, corresponding to nucleotides 2657–5812), the N-terminal region of the cylindrical inclusion to nuclear inclusion b protein coding region (CI) (CINIb, 1726 nt, corresponding to nucleotides 5577–7302), and the C-terminal region of the nuclear inclusion a (NIa) protein coding region to the 3′ NCR (NIa3NCR, 3019 nt, corresponding to nucleotides 6719–9737). The sequences of the RT-PCR products of adjacent regions of the genome overlapped by at least 200 bp. Sequences of each fragment were determined using three independent cloned plasmids. Each clone was sequenced by primer walking in both directions using a BigDye Terminator v3.1 Cycle Sequencing Ready Reaction kit (Applied Biosystems, USA) and an Applied Biosystems 3130 Genetic Analyzer. Sequence data were assembled using BioEdit version 5.0.9 (Hall, 1999).

2.3. Sequence alignment

Thirty sequences of each 5NCRP3, HC6K2, CINIB, and NIa3NCR region obtained in the present study (Table S1), which covers the complete genomic sequences, and 71 sequences of protein 1 (P1), 49 of HC-Pro, 51 of P3, 53 of NIa, and 126 of the CP coding region obtained from International nucleotide sequence databases were used for phylogenetic and evolutionary analyses (Table 1). TriMV (GenBank accession number, NC_012799) was used as the outgroup taxa because BLAST searches have shown it to be the sequence in the International nucleotide sequence databases most closely and consistently related to the genomic sequence of SCSMV (Fellers et al., 2009). The predicted SCSMV amino acid sequences were aligned with that of TriMV using CLUSTAL_X2 (Larkin et al., 2007) with TRANSALIGN (kindly supplied by Georg Weiller) to maintain the degapped alignment of the encoded amino acids. This produced sequence alignments for the 5NCRP3 (3214 nt), HC6K2 (3015 nt), CINIb (1422 nt) and NIa3NCR (2988 nt) regions, and the P1 (1071 nt), HC-Pro (1401 nt), P3 (894 nt), NIa (690 nt), and CP (831 nt) coding regions.

2.4. Recombination analyses

Putative recombination sites were identified using the RDP (Martin and Rybicki, 2000), GENECONV (Sawyer, 1999), BOOTSCAN (Salminen et al., 1995), MAXCHI (Smith, 1992), CHIMAERA (Posada and Crandall, 2001), 3SEQ (Boni et al., 2007), and SISCAN programs (Gibbs et al., 2000) as implemented in the RDP4 package (Martin et al., 2010), and also the original SISCAN version 2 program (Gibbs et al., 2000). These analyses were done using the default settings for the different detection programs and a Bonferroni corrected *P*-value cut-off of 0.05 or 0.01, and then isolates were identified as likely recombinants by the programs in RDP4 if they were supported by three different methods with an associated *P*-value of $<1.0 \times 10^{-6}$. These analyses also assessed which non-recombinant sequences had regions that were the closest to those from the recombinant sequences, indicating the likely lineages that provided those regions of the recombinant genomes. For convenience, we refer to them as the 'parental isolates' of the recombinants. Having examined all sites with an associated *P*-value of $<1 \times 10^{-6}$ (i.e., the most likely recombination sites), the intralineage recombinants (parents from the same major lineage) were retained, and the interlineage recombinants (parents from different major lineages) were removed by treating the identified recombination sites as missing data in subsequent analyses. The aligned 5′ and 3′ NCR sequences were added to the ends of the P1 and CP sequences, which were then assessed again for evidence of recombination, especially for recombination sites in the NCRs. Finally, SCSMV sequences were